

# A SUPPORT AND DENSITY THEOREM FOR MARKOVIAN ROUGH PATHS

ILYA CHEVYREV AND MARCEL OGRODNIK

**ABSTRACT.** We establish two results concerning a class of geometric rough paths  $\mathbf{X}$  which arise as Markov processes associated to uniformly subelliptic Dirichlet forms. The first is a support theorem for  $\mathbf{X}$  in  $\alpha$ -Hölder rough path topology for all  $\alpha \in (0, 1/2)$ , which answers in the positive a conjecture of Friz-Victoir [11]. The second is a Hörmander-type theorem for the existence of a density of a rough differential equation driven by  $\mathbf{X}$ , the proof of which is based on analysis of (non-symmetric) Dirichlet forms on manifolds.

## 1. INTRODUCTION

Consider a Markov process  $\mathbf{X}$  associated with a symmetric Dirichlet form on  $L^2(\mathbb{R}^d, \lambda)$

$$(1.1) \quad \mathcal{E}(f, g) = \int_{\mathbb{R}^d} \sum_{i,j=1}^d a^{i,j} (\partial_i f) (\partial_j g) d\lambda,$$

where  $\lambda$  is the Lebesgue measure, and  $a$  is a uniformly elliptic function taking values in symmetric  $d \times d$  matrices (we make our exact set-up precise in Section 1.1).

We are interested in differential equations of the form

$$(1.2) \quad d\mathbf{Y}_t = V(\mathbf{Y}_t) d\mathbf{X}_t, \quad \mathbf{Y}_0 = y_0 \in \mathbb{R}^e$$

driven by  $\mathbf{X}$  along vector fields  $V = (V_1, \dots, V_d)$  on  $\mathbb{R}^e$ . When  $a$  is taken sufficiently smooth, the process  $\mathbf{X}$  can be realised as a semi-martingale for which the classical framework of Itô gives meaning to the equation (1.2). However for irregular functions  $a$ , this is no longer the case, and (1.2) falls outside the scope of Itô calculus.

One of the applications of Lyons' theory of rough paths [16] has been to give meaning to differential equations driven by processes outside the range of semi-martingales. One viewpoint of rough paths theory is that it factors the problem of solving equations of the type (1.2) into first enhancing  $\mathbf{X}$  to a rough path by appropriately defining its iterated integrals (which is typically done through stochastic means), after which the theory allows (1.2) to be solved deterministically.

Probabilistic methods to enhance the Markov process  $\mathbf{X}$  to a rough path and the study of its fundamental properties appear in [15, 2, 13, 14], where primarily the forward-backward martingale decomposition is used to show existence of the stochastic area. A somewhat different approach, which we follow here, is taken in [9] where the authors define  $\mathbf{X}$  directly as a diffusion on the free nilpotent Lie group

---

2010 *Mathematics Subject Classification.* Primary 60H10; Secondary 60G17.

*Key words and phrases.* Markovian rough paths, support in Hölder topology, Hörmander's theorem.

$G^N(\mathbb{R}^d)$  (in particular the iterated integrals are given directly in the construction). One can show that in the situation mentioned at the start, the two methods give rise to equivalent definitions of rough paths, and the latter construction in fact yields further flexibility in that the evolution of  $\mathbf{X}$ , as a diffusion in  $G^N(\mathbb{R}^d)$ , can depend in a non-trivial way on its higher levels (its iterated integrals). Such Markovian rough paths have also recently been investigated in [6, 7] in connection with the accumulated local- $p$  variation functional and the moment problem for expected signatures.

The goal of this paper is to contribute two results to the study of Markovian rough paths in the sense of [9]. Our first contribution (Theorem 2.4) answers in the positive a conjecture about the support of  $\mathbf{X}$  in  $\alpha$ -Hölder rough path topology. Such a support theorem appeared in [9] for  $\alpha \in (0, 1/6)$ , and was improved to  $\alpha \in (0, 1/4)$  in [11] where it was conjectured to hold for  $\alpha \in (0, 1/2)$  in analogy to enhanced Brownian motion. Comparing our situation to the case of Gaussian rough paths, where such support theorems are known with sharp Hölder exponents (see e.g., [11] Section 15.8, and [10] for recent improvements), the difficulty of course lies in the lack of a Gaussian structure, in particular the absence of a Cameron-Martin space.

Our solution to this problem in fact uses few properties of  $\mathbf{X}$  beyond heat kernel estimates; indeed we first derive general conditions under which a stochastic process (taking values in a Polish space) admits explicit lower bounds on the probability of keeping a small  $\alpha$ -Hölder norm, and then verify these conditions for the translated (in general non-Markov) rough path  $T_h(\mathbf{X})$  for any  $h \in W^{1,2}([0, T], \mathbb{R}^d)$  (we also note that, just like for enhanced Brownian motion, all relevant constants depend on  $h$  only through  $\|h\|_{W^{1,2}}$ ). As usual, in combination with the continuity of the Itô-Lyons map from rough paths theory, an immediate consequence of improving the Hölder exponent in the support theorem for  $\mathbf{X}$  is a stronger Stroock-Varadhan support theorem (in  $\alpha$ -Hölder topology) for the solution  $\mathbf{Y}$  to the rough differential equation (RDE) (1.2) along with the lower regularity assumptions on the driving vector fields  $V$  ( $\text{Lip}^2$  instead of  $\text{Lip}^4$ ).

Our second contribution (Theorem 3.2 and its Corollary 3.4) may be seen as a non-Gaussian Hörmander-type theorem, and provides sufficient conditions on the driving vector fields  $V = (V_1, \dots, V_d)$  under which that the solution to the RDE (1.2) admits a density with respect to the Lebesgue measure on  $\mathbb{R}^e$ . Once again, while this result is reminiscent of density theorems for RDEs driven by Gaussian rough paths (e.g., [3, 4, 5]), the primary difference in our setting is that methods from Malliavin calculus are no longer available due to the lack of a Gaussian structure.

We replace the use of Malliavin calculus by direct analysis of (non-symmetric) Dirichlet forms on manifold. Indeed, we identify conditions under which the couple  $(\mathbf{X}, \mathbf{Y})$  admits a density on its natural state-space, and conclude by projecting to  $\mathbf{Y}$ . We note however that our current result gives no quantitative information about the density (beyond its existence) of even the couple  $(\mathbf{X}, \mathbf{Y})$ , and we strongly suspect that the method can be improved to yield further information (particularly  $L^p$  bounds and regularity results in the spirit of the De Giorgi–Nash–Moser theorem).

**1.1. Notation.** Throughout the paper, we adopt the convention that the domain of a path  $\mathbf{x} : [0, T] \mapsto E$ , for  $T > 0$  and a set  $E$ , is extended to all of  $[0, \infty)$  by setting  $\mathbf{x}_t = \mathbf{x}_T$  for all  $t > T$ .

We let  $G = G^N(\mathbb{R}^d)$  denote the step- $N$  free nilpotent Lie group over  $\mathbb{R}^d$  for some  $N \geq 2$ , and let  $U_1, \dots, U_d$  be a set of generators for its Lie algebra  $\mathfrak{g} = \mathfrak{g}^N(\mathbb{R}^d)$ , which we identify with the space of left-invariant vector fields on  $G$ . We equip  $\mathbb{R}^d$  with the inner product for which  $U_1, \dots, U_d$  form an orthonormal basis upon canonically identifying  $\mathbb{R}^d$  with a subspace of  $\mathfrak{g}$ .

For  $\Lambda > 0$ , let  $\Xi(\Lambda) = \Xi^{N,d}(\Lambda)$  denote the set of measurable functions  $a$  on  $G$  taking values in symmetric  $d \times d$  matrices which are sub-elliptic in the following sense:

$$\Lambda^{-1}|\xi|^2 \leq \langle \xi, a(x)\xi \rangle \leq \Lambda|\xi|^2, \quad \forall \xi \in \mathbb{R}^d, \quad \forall x \in G.$$

We let  $\lambda$  denote the Haar measure on  $G$  and define the associated Dirichlet form  $\mathcal{E} = \mathcal{E}^a$  on  $L^2(G, \lambda)$  for all  $f, g \in C_c^\infty(G)$  by

$$(1.3) \quad \mathcal{E}(f, g) = \int_G \sum_{i,j} a^{i,j}(U_i f)(U_j g) d\lambda.$$

We let  $\mathbf{X} = \mathbf{X}^{a,x}$  denote the Markov diffusion on  $G$  associated to  $\mathcal{E}$  with starting point  $\mathbf{X}_0 = x \in G$ . We recall that the sample paths of  $\mathbf{X}$  are a.s. geometric  $\alpha$ -Hölder rough paths for all  $\alpha \in (0, 1/2)$ , and when  $a(x)$  depends only on the level-1 projection  $\pi_1(x) \in \mathbb{R}^d$  of  $x \in G$ ,  $\mathbf{X}$  serves as the natural rough path lift of the Markov diffusion associated to the Dirichlet form (1.1) on  $L^2(\mathbb{R}^d)$  discussed earlier. For further details, we refer to [11].

*Remark 1.1.* Throughout the paper we assume the symmetric Dirichlet form (1.3) is defined on the Hilbert space  $L^2(G, \lambda)$  so that  $\mathbf{X}$  is symmetric with respect to  $\lambda$ . As pointed out in [6], it is natural to also consider  $\mathcal{E}$  defined over  $L^2(G, \mu)$  for a measure  $\mu(dx) = v(x)\lambda(dx)$ ,  $v \geq 0$ . While for simplicity we only work with  $\mathcal{E}$  defined on  $L^2(G, \lambda)$ , we note that appropriate assumptions of  $v$  and a Girsanov transform (see, e.g., [8]) can be used to relate the results of this paper to this more general setting.

**Acknowledgements.** I.C. is funded by a Junior Research Fellowship of St John's College, Oxford. M.O. is funded by a PhD scholarship from the Department of Mathematics, Imperial College London. The authors would like to thank Thomas Cass, Peter Friz, and Terry Lyons for helpful conversations on this topic.

## 2. SUPPORT THEOREM

**2.1. Restricted Hölder norms.** We first record some deterministic results on Hölder norms which will be used in the sequel. Throughout this section, let  $(E, d)$  be a metric space,  $\alpha \in (0, 1]$ ,  $T > 0$ , and  $\mathbf{x} \in C([0, T], E)$  a continuous path. Let  $\square$  denote any of the relations  $<, \leq, =, \geq, >$ , and consider the quantity

$$\|\mathbf{x}\|_{\alpha\text{-Höl}, \square; [s, t]} = \sup_{u, v \in [s, t], |u-v| \square \varepsilon} \frac{d(\mathbf{x}_u, \mathbf{x}_v)}{|u-v|^\alpha},$$

where we set  $\|\mathbf{x}\|_{\alpha\text{-Höl}, \square; [s, t]} = 0$  if the set  $\{(u, v) \in [s, t]^2 \mid |u-v| \square \varepsilon\}$  is empty.

For  $\varepsilon, \gamma > 0$  and  $s \in [0, T]$ , define also the times  $(\tau_n^{\varepsilon, \gamma, s})_{n \geq 0} = (\tau_n)_{n \geq 0}$  by  $\tau_0 = s$  and for  $n \geq 1$

$$\tau_n = \inf\{t > \tau_{n-1} \mid \|\mathbf{x}\|_{\alpha\text{-Höl}, \geq \varepsilon; [\tau_{n-1}, t]} \geq \gamma\}.$$

We call any such  $\tau_i$  a *Hölder stopping time* of  $\mathbf{x}$ .

**Lemma 2.1.** *Let  $\varepsilon, \gamma > 0$  and  $s = 0$ , and suppose that for some  $c > 0$*

$$(2.1) \quad \forall n \geq 0, \quad \sup_{t \in [\tau_n, \tau_n + \varepsilon]} d(\mathbf{x}_{\tau_n}, \mathbf{x}_t) < c.$$

*Then  $\|\mathbf{x}\|_{\alpha\text{-H\"{o}l},=\varepsilon;[0,T]} < \tilde{\gamma} := (3c\varepsilon^{-\alpha}) \vee (4\gamma + c\varepsilon^{-\alpha})$ .*

*Proof.* For  $n \geq 1$  and  $t \in [\tau_n - \varepsilon, \tau_n]$ , we have one of the following three mutually exclusive cases: (a)  $\tau_n = \tau_{n-1} + \varepsilon$ , (b)  $\tau_n \in (\tau_{n-1} + \varepsilon, \tau_{n-1} + 2\varepsilon]$  and  $t \in [\tau_{n-1}, \tau_{n-1} + \varepsilon]$ , or (c)  $t > \tau_{n-1} + \varepsilon$ . In case (a), (2.1) implies that  $d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) < 2c$ . In case (b),  $d(\mathbf{x}_{\tau_n}, \mathbf{x}_{\tau_{n-1}}) \leq \gamma(2\varepsilon)^\alpha$  and (2.1) implies that  $d(\mathbf{x}_t, \mathbf{x}_{\tau_{n-1}}) < c$ , so that

$$d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) < c + \gamma(2\varepsilon)^\alpha \leq (2c) \vee (4\gamma\varepsilon^\alpha).$$

In case (c), we have  $d(\mathbf{x}_{t-\varepsilon}, \mathbf{x}_t) < \gamma\varepsilon^\alpha$  and  $d(\mathbf{x}_{t-\varepsilon}, \mathbf{x}_{\tau_n}) \leq \gamma(2\varepsilon)^\alpha$ , so that

$$d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) < \gamma\varepsilon^\alpha + \gamma(2\varepsilon)^\alpha \leq 3\gamma\varepsilon^\alpha.$$

Hence, in all three cases

$$(2.2) \quad d(\mathbf{x}_t, \mathbf{x}_{\tau_n}) < (2c) \vee (4\gamma\varepsilon^\alpha).$$

Consider now

$$\tau = \inf\{t > 0 \mid \|\mathbf{x}\|_{\alpha\text{-H\"{o}l},=\varepsilon;[0,T]} = \tilde{\gamma}\}.$$

Note that  $\|\mathbf{x}\|_{\alpha\text{-H\"{o}l},=\varepsilon;[0,T]} \geq \tilde{\gamma} \Leftrightarrow \tau < \infty$ . Arguing by contradiction, suppose that  $\tau < \infty$ , which means that  $d(\mathbf{x}_{\tau-\varepsilon}, \mathbf{x}_\tau) = \tilde{\gamma}\varepsilon^\alpha$ . Consider the largest  $n$  for which  $\tau_n \leq \tau$ . Observe that  $\tau_n \in [\tau - \varepsilon, \tau]$ , since otherwise  $d(\mathbf{x}_{\tau-\varepsilon}, \mathbf{x}_\tau) < \gamma\varepsilon^\alpha$ , which is a contradiction since  $\tilde{\gamma} > \gamma$ . It follows from (2.1) that  $d(\mathbf{x}_{\tau_n}, \mathbf{x}_\tau) \leq c$ , and therefore by (2.2) and the triangle inequality

$$d(\mathbf{x}_{\tau-\varepsilon}, \mathbf{x}_\tau) < c + (2c) \vee (4\gamma\varepsilon^\alpha) = \tilde{\gamma}\varepsilon^\alpha,$$

which is again a contradiction.  $\square$

**Lemma 2.2.** *Suppose that  $\|\mathbf{x}\|_{\alpha\text{-H\"{o}l},=2^{-n}\varepsilon;[0,T]} \leq \gamma$  for every  $n > N \in \mathbb{Z}$ . Then*

$$\|\mathbf{x}\|_{\alpha\text{-H\"{o}l},<2^{-N}\varepsilon;[0,T]} \leq \frac{\gamma}{1 - 2^{-\alpha}}.$$

*Proof.* Consider  $(t - s)/\varepsilon \in (0, 2^{-N})$  with the binary representation  $(t - s)/\varepsilon = \sum_{n=m}^{\infty} c_n 2^{-n}$  with  $c_n \in \{0, 1\}$ ,  $m > N$ , and  $c_m = 1$ . It follows that

$$d(\mathbf{x}_s, \mathbf{x}_t) \leq \gamma \sum_{n=m}^{\infty} \varepsilon^\alpha c_n 2^{-n\alpha}.$$

Since  $2^{-m} \leq (t - s)/\varepsilon$ , we have  $\varepsilon^\alpha 2^{-n\alpha} \leq 2^{\alpha(m-n)}(t - s)^\alpha$ . Hence

$$d(\mathbf{x}_s, \mathbf{x}_t) \leq \gamma \sum_{n=m}^{\infty} 2^{\alpha(m-n)}(t - s)^\alpha = \frac{\gamma(t - s)^\alpha}{1 - 2^{-\alpha}}.$$

$\square$

**Lemma 2.3.** *Suppose there exist  $x \in E$  and  $r > 0$  such that for all integers  $k \geq 0$ ,  $\mathbf{x}_{k\varepsilon} \in B(x, r)$  and  $\|\mathbf{x}\|_{\alpha\text{-H\"{o}l},\leq\varepsilon;[k\varepsilon, (k+1)\varepsilon]} \leq \gamma$ . Then*

$$\|\mathbf{x}\|_{\alpha\text{-H\"{o}l},[0,T]} \leq 2\gamma + 2r\varepsilon^{-\alpha}.$$

*Proof.* Consider  $0 \leq s < t \leq [0, T]$ , and denote  $s \in [k\varepsilon, (k+1)\varepsilon)$ ,  $t \in [n\varepsilon, (n+1)\varepsilon)$ . If  $k = n$  there is nothing to prove, so suppose  $k < n$ . If  $|t - s| \leq \varepsilon$ , so that  $n = k+1$ , then

$$d(\mathbf{x}_s, \mathbf{x}_t) \leq d(\mathbf{x}_s, \mathbf{x}_{n\varepsilon}) + d(\mathbf{x}_{n\varepsilon}, \mathbf{x}_t) \leq \gamma 2^{1-\alpha} |t - s|^\alpha.$$

Finally, if  $|t - s| > \varepsilon$  then since  $\mathbf{x}_{k\varepsilon}, \mathbf{x}_{n\varepsilon} \in B(x, r)$ , it follows that

$$\begin{aligned} |t - s|^{-\alpha} d(\mathbf{x}_s, \mathbf{x}_t) &\leq |t - s|^{-\alpha} (d(\mathbf{x}_{k\varepsilon}, \mathbf{x}_s) + d(\mathbf{x}_{k\varepsilon}, \mathbf{x}_{n\varepsilon}) + d(\mathbf{x}_{n\varepsilon}, \mathbf{x}_t)) \\ &\leq |t - s|^{-\alpha} (2\varepsilon^\alpha \gamma + 2r) \\ &\leq 2\gamma + 2r\varepsilon^{-\alpha}. \end{aligned}$$

□

**2.2. Positive probability of small Hölder norm.** Suppose now  $(E, d)$  is a Polish space. In this section, we give conditions under which an  $E$ -valued process has an explicit positive probability of keeping a small Hölder norm. We fix  $\alpha \in (0, 1/2)$ , a terminal time  $T > 0$ , and an  $E$ -valued stochastic process  $\mathbf{X}$  adapted to a filtration  $(\mathcal{F}_t)_{t \in [0, T]}$ .

Consider the following conditions:

- (1) There exists  $C_1 > 0$  such that for every  $c, \varepsilon > 0$ , and every Hölder stopping time  $\tau$  of  $\mathbf{X}$ , a.s.

$$\mathbb{P} \left[ \sup_{t \in [\tau, \tau + \varepsilon]} d(\mathbf{X}_\tau, \mathbf{X}_t) > c \mid \mathcal{F}_\tau \right] \leq C_1 \exp \left( \frac{-c^2}{C_1 \varepsilon} \right).$$

- (2) There exist  $c_2, C_2 > 0$  and  $x \in E$  such that for every  $s \in [0, T]$  and  $\varepsilon \in (0, T - s]$ , a.s.

$$\mathbb{P} \left[ \mathbf{X}_{s+\varepsilon} \in B(x, C_2 \varepsilon^{1/2}) \mid \mathcal{F}_s \right] \geq c_2 \mathbf{1}\{\mathbf{X}_s \in B(x, C_2 \varepsilon^{1/2})\}.$$

Roughly speaking, the first condition states that the probability of large fluctuations of  $\mathbf{X}$  over small time intervals should have the same Gaussian tails as that of a Brownian motion, while the second condition bounds from below the probability that  $\mathbf{X}_{s+\varepsilon}$  is in a ball of radius  $\sim \varepsilon^{1/2}$  given that  $\mathbf{X}_s$  was in the same ball.

**Theorem 2.4.** *Assume conditions (1) and (2). Then there exist  $C_{2.4}, c_{2.4} > 0$ , depending only on  $C_1, c_2, C_2, \alpha$  and  $T$ , such that for every  $\gamma > 0$ , a.s.*

$$\mathbb{P} \left[ \|\mathbf{X}\|_{\alpha\text{-Hölder}; [0, T]} < \gamma \mid \mathcal{F}_0 \right] \geq C_{2.4}^{-1} \exp \left( \frac{-C_{2.4}}{\gamma^{2/(1-2\alpha)}} \right) \mathbf{1}\{\mathbf{X}_0 \in B(x, c_{2.4} \gamma^{1/(1-2\alpha)})\}.$$

**Lemma 2.5.** *Assume condition (1). Then there exists  $C_{2.5} > 0$ , depending only on  $C_1$  and  $\alpha$ , such that for all  $0 \leq s < t \leq T$  and  $\varepsilon \in (0, t - s]$ , a.s.*

$$\mathbb{P} \left[ \|\mathbf{X}\|_{\alpha\text{-Hölder}; [s, t]} \geq \gamma \mid \mathcal{F}_s \right] \leq C_{2.5} (t-s) \varepsilon^{-1} (\gamma^{-2} \varepsilon^{1-2\alpha} + 1) \exp \left( \frac{-\gamma^2 (1 - 2^{-\alpha})^2}{9C_1 \varepsilon^{1-2\alpha}} \right).$$

*Proof.* Let  $\tau_n = \tau_n^{\varepsilon, \gamma, s}$  be defined as usual with  $\tau_0 = s$ . Note that (1) implies that for all  $c, \gamma > 0$ ,  $t > s$  and  $\varepsilon \in (0, t - s]$ ,

$$\mathbb{P} \left[ \exists n \geq 0, \tau_n \leq t, \sup_{u \in [\tau_n, \tau_n + \varepsilon]} d(\mathbf{X}_{\tau_n}, \mathbf{X}_u) > c \mid \mathcal{F}_s \right] \leq \lceil (t-s)/\varepsilon \rceil C_1 \exp \left( \frac{-c^2}{C_1 \varepsilon} \right),$$

so that by Lemma 2.1

$$\mathbb{P} \left[ \|\mathbf{X}\|_{\alpha\text{-H\"ol};=\varepsilon;[s,t]} \geq (3c\varepsilon^{-\alpha}) \vee (4\gamma + c\varepsilon^{-\alpha}) \mid \mathcal{F}_s \right] \leq \lceil (t-s)/\varepsilon \rceil C_1 \exp \left( \frac{-c^2}{C_1\varepsilon} \right).$$

In particular, choosing  $c = 2\gamma\varepsilon^\alpha$  yields that for all  $\gamma > 0$ ,  $t > s$ , and  $\varepsilon \in (0, t-s]$ ,

$$\mathbb{P} \left[ \|\mathbf{X}\|_{\alpha\text{-H\"ol};=\varepsilon;[s,t]} \geq 6\gamma \mid \mathcal{F}_s \right] \leq \lceil (t-s)/\varepsilon \rceil C_1 \exp \left( \frac{-(2\gamma)^2}{C_1\varepsilon^{1-2\alpha}} \right).$$

Hence

$$\mathbb{P} \left[ \exists n \geq 0, \|\mathbf{X}\|_{\alpha\text{-H\"ol};=2^{-n}\varepsilon;[s,t]} \geq \gamma \mid \mathcal{F}_s \right] \leq 2C_1(t-s)\varepsilon^{-1} \sum_{n=0}^{\infty} 2^n \exp \left( \frac{-2^{n(1-2\alpha)}\gamma^2}{9C_1\varepsilon^{1-2\alpha}} \right).$$

The conclusion now follows from Lemma 2.2 and the observation that for every  $\theta > 0$  there exists  $C_4$  such that for all  $K > 0$

$$\sum_{n=0}^{\infty} 2^n \exp(-K2^{\theta n}) \leq C_4(K^{-1} + 1)e^{-K}$$

(which can be seen, for example, by the integral test and the asymptotic behaviour of the incomplete gamma function  $\Gamma(p, K)$ ).  $\square$

*Proof of Theorem 2.4.* For  $\gamma, \varepsilon > 0$  and  $s \in [0, T]$ , consider the event

$$A_s = \{ \|\mathbf{X}\|_{\alpha\text{-H\"ol};\leq\varepsilon;[s,s+\varepsilon]} < \gamma, \mathbf{X}_{s+\varepsilon} \in B(x, C_2\varepsilon^{1/2}) \}.$$

Applying condition (2) and Lemma 2.5 with  $t = s + \varepsilon$ , we see that for all  $s \in [0, T]$ , and  $\varepsilon, \gamma > 0$

$$\mathbb{P}[A_s \mid \mathcal{F}_s] \geq c_2 \mathbf{1}\{\mathbf{X}_s \in B(x, C_2\varepsilon^{1/2})\} - C_{2.5}(\gamma^{-2}\varepsilon^{1-2\alpha} + 1) \exp \left( \frac{-\gamma^2(1-2^{-\alpha})^2}{9C_1\varepsilon^{1-2\alpha}} \right).$$

Observe also that Lemma 2.3 (with  $r = C_2\varepsilon^{1/2}$ ) implies that for all  $\varepsilon, \gamma > 0$

$$\mathbb{P} \left[ \|\mathbf{X}\|_{\alpha\text{-H\"ol};[0,T]} < 2\gamma + 2C_2\varepsilon^{1/2-\alpha} \mid \mathcal{F}_0 \right] \geq \mathbb{P} \left[ \bigcap_{k=0}^{\lceil T/\varepsilon \rceil - 1} A_{k\varepsilon} \mid \mathcal{F}_0 \right].$$

It remains to control the final last probability on the RHS. We set  $\varepsilon = c_1\gamma^{2/(1-2\alpha)}$  (so that  $\varepsilon^{1/2-\alpha} \sim \gamma$ ), where  $c_1 > 0$  is sufficiently small (and depends only on  $C_1, c_2, C_2, C_{2.5}$  and  $\alpha$ ) such that

$$\kappa := c_2 - C_{2.5}(c_1^{1-2\alpha} + 1) \exp \left( \frac{-(1-2^{-\alpha})^2}{36C_1c_1^{1-2\alpha}} \right) > 0.$$

so in particular for all  $s \in [0, T]$  and  $\gamma > 0$ ,

$$\mathbb{P}[A_s \mid \mathcal{F}_s] \geq \kappa \mathbf{1}\{\mathbf{X}_s \in B(x, C_2\varepsilon^{1/2})\}.$$

Inductively applying conditional expectations, it follows that for all  $n \geq 0$

$$\mathbb{P} \left[ \bigcap_{k=0}^n A_{k\varepsilon} \mid \mathcal{F}_0 \right] \geq \kappa^{n+1} \mathbf{1}\{\mathbf{X}_0 \in B(x, C_2\varepsilon^{1/2})\}.$$

Taking  $n = \lceil T/\varepsilon \rceil - 1$  yields the desired result.  $\square$

**2.3. Support theorem for Markovian rough paths.** We now turn to the support theorem for Markovian rough paths in  $\alpha$ -Hölder topology. Recall Notation 1.1 and equip  $G = G^N(\mathbb{R}^d)$  with the Carnot-Carathéodory metric  $d$ .

Recall the Sobolev path space  $W^{1,2} = W^{1,2}([0, T], \mathbb{R}^d)$  and the translation operator  $T_h(\mathbf{x})$  defined for  $\mathbf{x} \in C^{p\text{-var}}([0, T], G)$ ,  $1 \leq p < N+1$ , and  $h \in C^{1\text{-var}}([0, T], \mathbb{R}^d)$  (see [11] Section 9.4).

Let us fix  $\alpha \in (0, 1/2)$ . We will show that for all  $h \in W^{1,2}$  and  $\gamma > 0$ ,

$$\mathbb{P}^{a,x} [d_{\alpha\text{-Hölder};[0,T]}(\mathbf{X}, S_N(h)) < \gamma] > 0,$$

where  $d_{\alpha\text{-Hölder};[0,T]}$  denotes the (homogeneous)  $\alpha$ -Hölder metric. Indeed, this is an immediate consequence of the continuity of  $T_h(\cdot)$  in  $d_{\alpha\text{-Hölder};[0,T]}$  and of the following result.

**Theorem 2.6.** *Let  $h \in W^{1,2}$ . There exists a constant  $C_{2.6}$ , depending only on  $\Lambda$ ,  $\|h\|_{W^{1,2}}$ ,  $\alpha$ , and  $T$ , such that for all  $a \in \Xi(\Lambda)$ ,  $x \in G$  and  $\gamma > 0$*

$$\mathbb{P}^{a,x} [\|T_h(\mathbf{X})\|_{\alpha\text{-Hölder};[0,T]} < \gamma] \geq C_{2.6}^{-1} \exp\left(\frac{-C_{2.6}}{\gamma^{2/(1-2\alpha)}}\right).$$

For the proof, recall that the Fernique estimate ([11] Corollary 16.12) implies that for every stopping time  $\tau$  and  $p > 2$ , a.s.

$$(2.3) \quad \mathbb{P} [\|\mathbf{X}\|_{p\text{-var};[\tau, \tau+\varepsilon]} > c \mid \mathcal{F}_\tau] \leq C_F \exp\left(\frac{-c^2}{C_F \varepsilon}\right),$$

where  $C_F$  depends only on  $\Lambda$  and  $p$ . Moreover, recall that

$$\|h\|_{1\text{-var};[s, s+\varepsilon]} \leq C_W \varepsilon^{1/2} \|h\|_{W^{1,2};[s, s+\varepsilon]}$$

for a universal constant  $C_W$ . We now prove two lemmas which demonstrate that the (in general non-Markov) process  $T_h(\mathbf{X})$  satisfies conditions (1) and (2).

**Lemma 2.7.** *There exists a constant  $C$ , depending only on  $\Lambda$ , such that for all  $c, \varepsilon > 0$  satisfying*

$$(2.4) \quad \varepsilon \leq \frac{c^2}{4C_W^2 \|h\|_{W^{1,2}}^2},$$

*it holds that for every stopping time  $\tau$ , a.s.*

$$\mathbb{P} \left[ \sup_{t \in [\tau, \tau+\varepsilon]} d(T_h(\mathbf{X})_\tau, T_h(\mathbf{X})_t) > c \mid \mathcal{F}_\tau \right] \leq C \exp\left(\frac{-c^2}{4C\varepsilon}\right).$$

*Proof.* Fix any  $2 < p < N+1$ . Note that (see, e.g., [11] Exercise 9.37)

$$\begin{aligned} \sup_{t \in [s, s+\varepsilon]} d(T_h(\mathbf{X})_s, T_h(\mathbf{X})_t) &\leq \|T_h(\mathbf{X})\|_{p\text{-var};[s, s+\varepsilon]} \\ &\leq C_1 \left( \|\mathbf{X}\|_{p\text{-var};[s, s+\varepsilon]} + \|h\|_{1\text{-var};[s, s+\varepsilon]} \right). \end{aligned}$$

Hence whenever  $c, \varepsilon > 0$  satisfy (2.4), we have  $\|h\|_{1\text{-var};[s, s+\varepsilon]} \leq c/2$ , and the conclusion follows from the Fernique estimate (2.3).  $\square$

**Lemma 2.8.** *For all  $C \geq C_0(\Lambda, \|h\|_{W^{1,2}}) > 0$ , there exists  $c = c(C, \Lambda, \|h\|_{W^{1,2}}) > 0$  such that for all  $x \in G$ ,  $s \in [0, T]$ , and  $\varepsilon \in (0, T-s]$ , a.s.*

$$\mathbb{P} [T_h(\mathbf{X})_{s+\varepsilon} \in B(x, C\varepsilon^{1/2}) \mid \mathcal{F}_s] \geq c \mathbf{1}\{T_h(\mathbf{X})_s \in B(x, C\varepsilon^{1/2})\}.$$

*Proof.* We use the shorthand notation  $\mathbf{Y} = T_h(\mathbf{X})$ . For every  $x, y \in G$ , consider a geodesic  $\gamma^{y,x} : [0, 1] \mapsto G$  with  $\gamma_0^{y,x} = y$  and  $\gamma_1^{y,x} = x$  parametrised at unit speed. Let  $z(y, x) := \gamma_{1/2}^{y,x}$  denote its midpoint. For any  $x \in G$ , observe that

$$\begin{aligned} d(\mathbf{Y}_{s+\varepsilon}, x) &\leq d(\mathbf{Y}_{s+\varepsilon}, z(\mathbf{Y}_s, x)) + d(z(\mathbf{Y}_s, x), x) \\ &\leq d(\mathbf{Y}_{s+\varepsilon}, \mathbf{X}_{s,s+\varepsilon}) + d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_s^{-1} z(\mathbf{Y}_s, x)) + d(z(\mathbf{Y}_s, x), x). \end{aligned}$$

If  $\mathbf{Y}_s \in B(x, r)$ , then evidently  $d(z(\mathbf{Y}_s, x), x) \leq r/2$ . Moreover, by lower bounds on the heat kernel and the volume of balls ([11] Chapter 16), there exists  $C_1 > 0$ , depending only on  $\Lambda$ , such that for all  $x \in G$ ,  $r, \varepsilon > 0$  and  $s \in [0, T]$

$$\mathbb{P}[d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_s^{-1} z(\mathbf{Y}_s, x)) < r/4 \mid \mathcal{F}_s] \geq C_1^{-1} \exp\left(\frac{-C_1 r^2}{\varepsilon}\right) \mathbf{1}\{\mathbf{Y}_s \in B(x, r)\}.$$

Finally, by standard rough paths estimates (using that  $T_h(\mathbf{X})_{s,t}$  is equal to  $\mathbf{X}_{s,t}$  plus a combination cross-integrals of  $\mathbf{X}$  and  $h$  over  $[s, t]$ ) we have

$$\begin{aligned} d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_{s,s+\varepsilon}) &\leq C_2 \max_{i \in \{1, \dots, N\}} \left( \sum_{k=1}^i \|h\|_{1\text{-var};[s,s+\varepsilon]}^k \|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]}^{i-k} \right)^{1/i} \\ &\leq C_3 \max_{i \in \{1, \dots, N\}} \left( \sum_{k=1}^i \varepsilon^{k/2} \|h\|_{W^{1,2};[s,s+\varepsilon]}^k \|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]}^{i-k} \right)^{1/i}. \end{aligned}$$

Hence, if  $\|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]} \leq R\varepsilon^{1/2}$ , then for some  $C_4 > 0$  depending only on  $G$

$$d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_{s,s+\varepsilon}) \leq C_4 \varepsilon^{1/2} (\|h\|_{W^{1,2};[s,s+\varepsilon]}^{1/N} + \|h\|_{W^{1,2};[s,s+\varepsilon]})(1 + R^{(N-1)/N}).$$

We now let  $r = C\varepsilon^{1/2}$ . It follows that if  $C$  and  $R$  satisfy

$$(2.5) \quad C \geq 4C_4 (\|h\|_{W^{1,2};[s,s+\varepsilon]}^{1/N} + \|h\|_{W^{1,2};[s,s+\varepsilon]})(1 + R^{(N-1)/N}),$$

then by the Fernique estimate (2.3), for any  $2 < p < N + 1$ ,

$$\begin{aligned} \mathbb{P}[d(\mathbf{X}_{s,s+\varepsilon}, \mathbf{Y}_{s,s+\varepsilon}) > C\varepsilon^{1/2}/4 \mid \mathcal{F}_s] &\leq \mathbb{P}[\|\mathbf{X}\|_{p\text{-var};[s,s+\varepsilon]} > R\varepsilon^{1/2}, \mid \mathcal{F}_s] \\ &\leq C_F \exp\left(\frac{-R^2}{C_F}\right). \end{aligned}$$

It follows that if  $C$  and  $R$  furthermore satisfy

$$(2.6) \quad c := C_1^{-1} \exp(-C_1 C^2) - C_F \exp\left(\frac{-R^2}{C_F}\right) > 0$$

then we obtain

$$\mathbb{P}[d(\mathbf{Y}_{s+\varepsilon}, x) < C\varepsilon^{1/2} \mid \mathcal{F}_s] \geq c \mathbf{1}\{\mathbf{Y}_s \in B(x, C\varepsilon^{1/2})\}.$$

We now observe that due to the factor  $R^{(N-1)/N}$  in (2.5) above, there exists  $C_0 > 0$ , depending only on  $\|h\|_{W^{1,2}}$  and  $\Lambda$ , such that for every  $C \geq C_0$ , we can find  $R > 0$  for which (2.5) and (2.6) are satisfied.  $\square$

*Proof of Theorem 2.6.* By Theorem 2.4, it suffices to check that  $T_h(\mathbf{X})$  satisfies conditions (1) and (2) with constants  $C_1, c_2, C_2$  only depending on  $\Lambda$  and  $\|h\|_{W^{1,2}}$ . However this follows directly from Lemmas 2.7 and 2.8.  $\square$



## 3. DENSITY THEOREM

**3.1. Semi-Dirichlet associated with Hörmander vector fields.** In this subsection, let  $\mathcal{O}$  be a smooth manifold and  $W = (W_1, \dots, W_d)$  a collection of smooth vector fields on  $\mathcal{O}$ . For  $z \in \mathcal{O}$ , let  $\text{Lie}_z W$  denote the subspace of  $T_z \mathcal{O}$  spanned by the vector fields  $(W_1, \dots, W_d)$  and all their commutators at  $z$ . We say that  $W$  satisfies Hörmander's condition on  $\mathcal{O}$  if  $\text{Lie}_z W = T_z \mathcal{O}$  for every  $z \in \mathcal{O}$ , in which case we call  $W$  a collection of Hörmander vector fields.

Fix a collection  $W = (W_1, \dots, W_d)$  of Hörmander vector fields on  $\mathcal{O}$  and  $U \subset \mathcal{O}$  an open subset with compact closure. Consider a bounded measurable function  $a$  on  $U$  taking values in (not necessarily symmetric)  $d \times d$  matrices such that for some  $\Lambda \geq 1$

$$(3.1) \quad \Lambda^{-1}|\xi|^2 \leq \langle \xi, a(z)\xi \rangle, \quad \forall \xi \in \mathbb{R}^d, \quad \forall z \in U.$$

Let  $\mu$  be a smooth measure on  $\mathcal{O}$  and define the bilinear map

$$\begin{aligned} \mathcal{E} : C_c^\infty(U) \times C_c^\infty(U) &\mapsto \mathbb{R} \\ \mathcal{E} : (f, g) &\mapsto - \sum_{i,j=1}^d \int_U a^{i,j}(z) (W_i f)(z) (W_j^* g)(z) \mu(dz), \end{aligned}$$

where  $W_j^* = -W_j - \text{div}_\mu W_j$  is the formal adjoint of  $W_j$  with respect to  $\mu$ . The following is the main result of this subsection. For background concerning (non-symmetric, semi-)Dirichlet forms, we refer to [18]. The  $L^p$  norm  $\|\cdot\|_p$  for  $p \in [1, \infty]$  is assumed to be on  $L^p(U, \mu)$ .

**Proposition 3.1.** *The bilinear form  $\mathcal{E}$  is closable in  $L^2(U, \mu)$ , lower bounded, and satisfies the sector condition.*

Let  $P_t$  denote the associated (strongly continuous) semi-group on  $L^2(U, \mu)$ . Suppose further that  $P_t$  is sub-Markov (so that the closed extension of  $\mathcal{E}$  is a lower-bounded semi-Dirichlet forms) and maps  $C_b(U)$  into itself. Then there exists  $\nu > 2$  and  $b > 0$  such that for every  $x \in U$  and  $t > 0$  there exists  $p_t(x, \cdot) \in L^2(U, \mu)$  with  $\|p_t(x, \cdot)\|_2 \leq bt^{-\nu/2}$  such that for all  $f \in L^2(U, \mu)$

$$P_t f(x) = \int_U p_t(x, y) f(y) \mu(dy).$$

The proof of Proposition 3.1, which we defer to Appendix A, is based on the sub-Riemannian Sobolev inequality combined with a classical argument of Nash [17].

**3.2. Density for RDEs.** We now specialise to the setting of Markovian rough paths. Recall Notation 1.1 and consider the RDE

$$(3.2) \quad d\mathbf{Y}_t = V(\mathbf{Y}_t) d\mathbf{X}_t, \quad \mathbf{Y}_0 = y_0 \in \mathbb{R}^e,$$

for smooth vector fields  $V = (V_1, \dots, V_d)$  on  $\mathbb{R}^e$ . We suppose also that  $V$  are  $\text{Lip}^2$  so that (3.2) admits a unique solution. We fix also the starting point  $\mathbf{X}_0 = x_0 \in G$  of  $\mathbf{X}$ .

Consider the manifold  $G \times \mathbb{R}^e$ . We canonically identify the tangent space  $T_{(x,y)}(G \times \mathbb{R}^e)$  with  $T_x G \oplus T_y \mathbb{R}^e$  and define smooth vector fields on  $G \times \mathbb{R}^e$  by  $W_i = U_i + V_i$ . Let  $z_0 = (x_0, y_0) \in G \times \mathbb{R}^e$  and denote by  $\mathcal{O} = \mathcal{O}_{z_0}$  the orbit of  $z_0$  under the collection  $W = (W_1, \dots, W_d)$ . By the orbit theorem (e.g., [1] Chapter 5),  $\mathcal{O}$  is an immersed submanifold of  $G \times \mathbb{R}^e$  and  $\text{Lie}_z W \subseteq T_z \mathcal{O}$  for all  $z \in \mathcal{O}$ .

Denote the couple  $\mathbf{Z}_t = (\mathbf{X}_t, \mathbf{Y}_t)$  which is a Markov process on  $G \times \mathbb{R}^e$ . One can readily show that a.s.  $\mathbf{Z}_t^{z_0} \in \mathcal{O}$  for all  $t > 0$  (e.g., by approximating each sample path of  $\mathbf{X}$  in  $p$ -variation for some  $p > 2$  by piecewise linear paths).

**Theorem 3.2.** *Suppose  $W$  satisfies Hörmander's condition on  $\mathcal{O}$ , i.e.,  $\text{Lie}_z W = T_z \mathcal{O}$  for all  $z \in \mathcal{O}$ . Then for all  $t > 0$ ,  $\mathbf{Z}_t^{z_0}$  admits a density with respect to any smooth measure  $\mu$  on  $\mathcal{O}$ .*

*Remark 3.3.* In the special case that  $a(x)$  depends only on the first level  $\pi_1(x)$  for all  $x \in G^2(\mathbb{R}^d)$ , the identical statement in Theorem 3.2 holds for the process  $\mathbf{Z}_t = (\pi_1(\mathbf{X}_t), \mathbf{Y}_t)$  upon setting  $N = 1$  and  $G = G^1(\mathbb{R}^d) \cong \mathbb{R}^d$ , and going through the same discussion above. The reason for this is that Lemma 3.8 below can be readily adjusted to give analogous infinitesimal behaviour of the process  $\mathbf{Z}_t$  (now taking values in  $\mathcal{O} \subseteq \mathbb{R}^d \times \mathbb{R}^e$ ), after which the proof of the theorem carries through without change.

For a statement of the density of  $\mathbf{Y}_t$  itself, let  $\mathcal{O}' \subseteq \mathbb{R}^e$  denote the orbit of  $y_0 \in \mathbb{R}^e$  under  $V$  (e.g.,  $\mathcal{O}' = \mathbb{R}^e$  whenever  $V$  satisfies Hörmander's condition on  $\mathbb{R}^e$ ). By the description of the tangent space  $T_z \mathcal{O}$  in the orbit theorem, it holds that the projection  $p_2 : \mathcal{O} \mapsto \mathcal{O}'$ ,  $(x, y) \mapsto y$ , is a (surjective) submersion (in fact a smooth fibre bundle) from  $\mathcal{O}$  to  $\mathcal{O}'$ . Since pre-images of null-sets under submersions are null-sets, we see that  $\mathbf{Y}_t$  admits a density in  $\mathcal{O}'$  whenever  $\mathbf{Z}_t^{z_0}$  admits a density in  $\mathcal{O}$ .

Furthermore, the condition in the theorem may be easily restated in terms of just the driving vector fields  $V = (V_1, \dots, V_d)$ . Indeed, by the Frobenius theorem and the free step- $N$  nilpotency of  $G$ , it holds that  $W$  satisfies Hörmander's condition on  $\mathcal{O}$  if and only if

$$(3.3) \quad \text{the dimension of } \text{span}\{V_{[I]}(y) : |I| > N\} \subseteq T_y \mathbb{R}^e \text{ is constant in } y \in \mathcal{O}'.$$

In the above, we denote by  $V_{[I]}$  the vector field  $[[\dots [V_{i_1}, V_{i_2}], \dots], V_{i_k}]$  for a multi-index  $I = (i_1, \dots, i_k) \in \{1, \dots, d\}^k$  of length  $|I| = k$ .

We thus obtain the following corollary of Theorem 3.2.

**Corollary 3.4.** *Suppose condition (3.3) holds. Then for all  $t > 0$ , the RDE solution  $\mathbf{Y}_t$  admits a density with respect to any smooth measure on  $\mathcal{O}'$ .*

*Remark 3.5.* Following Remark 3.3, in the case that  $a(x)$  depends only on the first level  $\pi_1(x)$ , we are able to take  $N = 1$  in (3.3) when applying Corollary 3.4.

*Remark 3.6.* Note that while (3.3) (for any  $N \geq 0$ ) implies that  $V$  satisfies Hörmander's condition on  $\mathcal{O}'$ , the reverse implication is clearly not true. In particular, we do not know if it is sufficient for  $V$  to only satisfy Hörmander's condition on  $\mathcal{O}'$  in order for  $\mathbf{Y}_t$  to admit a density on  $\mathcal{O}'$ . The difficulty of course is that unless (3.3) is satisfied, the couple  $(\mathbf{X}_t, \mathbf{Y}_t)$  will in general not admit a density in  $\mathcal{O}$ , whereby our method of proof breaks down.

For the proof of Theorem 3.2, we first recall for the reader's convenience the infinitesimal behaviour of the coordinate projections of  $\mathbf{X}^a$ . As before, let  $\lambda$  denote the Haar measure on  $G$ .

**Lemma 3.7.** *Let  $g \in C_c^\infty(G)$ . Then for all  $k, l \in \{1, \dots, d\}$*

$$\begin{aligned} \lim_{t \rightarrow 0} t^{-1} \langle g, \mathbb{E}^{a, \cdot} [\mathbf{X}_{0,t}^k] \rangle_{L^2(G, \lambda)} &= - \sum_{j=1}^d \int_G a^{k,j}(x) U_j g(x) \lambda(dx), \\ \lim_{t \rightarrow 0} t^{-1} \langle g, \mathbb{E}^{a, \cdot} [\mathbf{X}_{0,t}^k \mathbf{X}_{0,t}^l] \rangle_{L^2(G, \lambda)} &= 2 \int_G a^{k,l}(x) g(x) \lambda(dx), \\ \lim_{t \rightarrow 0} t^{-1} \langle g, \mathbb{E}^{a, \cdot} [\mathbf{X}_{0,t}^{k,l}] \rangle_{L^2(G, \lambda)} &= 0. \end{aligned}$$

*Proof.* This is [9] Lemma 27 extended *mutatis mutandis* to the general case  $G^N(\mathbb{R}^d)$ ,  $N \geq 1$ , cf. [11] Proposition 16.20.  $\square$

**Lemma 3.8.** *Let  $U \subset \mathcal{O}$  be an open subset with compact closure. Consider the (sub-Markov) semi-group  $P_t^U$  of  $\mathbf{Z}_t$  killed upon exiting  $U$ , defined for all bounded measurable  $f : U \mapsto \mathbb{R}$  by*

$$P_t^U f(z) = \mathbb{E}^z [f(\mathbf{Z}_t) \mathbf{1}\{\mathbf{Z}_s \in U, \forall s \in [0, t]\}].$$

*Then  $P_t^U$  maps  $C_b(U)$  into itself, and for any smooth measure  $\mu$  on  $\mathcal{O}$  it holds that for all  $f, g \in C_c^\infty(U)$*

$$(3.4) \quad \lim_{t \rightarrow 0} t^{-1} \langle P_t^U f - f, g \rangle_{L^2(U, \mu)} = \sum_{i,j=1}^d \int_U a^{i,j}(p_1(z)) (W_i f)(z) (W_j^* g)(z) \mu(dz),$$

where  $W_j^* = -W_j - \text{div}_\mu(W_j)$  is the adjoint of  $W_j$  in  $L^2(U, \mu)$ .

*Proof.* To show that  $P_t^U$  map  $C_b(U)$  into itself, let  $f \in C_b(U)$ . As  $z_n = (x_n, y_n) \rightarrow z = (x, y)$  in  $U$ , it holds in particular that  $x_n \rightarrow x$  in  $G$ . It follows that  $\mathbf{X}^{a, x_n} \xrightarrow{\mathcal{D}} \mathbf{X}^{a, x}$  in  $\alpha$ -Hölder topology for any  $\alpha \in [0, 1/2)$  (Theorem 16.28 [11]) and we readily obtain that  $P_t^U f(z_n) \rightarrow P_t^U f(z)$ . Hence  $P_t^U f \in C_b(U)$ , so indeed  $P_t^U$  map  $C_b(U)$  into itself.

It remains to verify (3.4). Note that for every  $z \in U$  the probability that  $\mathbf{Z}^z$  leaves  $U$  in  $[0, t]$  is bounded above by  $C^{-1} \exp(-Ct^{-1})$  for some  $C = C(z, U, \Lambda) > 0$  (see, e.g., the Fernique estimate (2.3)). It follows by a localisation argument and the stochastic Taylor expansion (e.g., [9] Lemma 26), that

$$\begin{aligned} \lim_{t \rightarrow 0} t^{-1} \langle P_t^U f - f, g \rangle_{L^2(U, \mu)} &= \lim_{t \rightarrow 0} t^{-1} \int_U \left( \sum_{i=1}^d W_i f(z) \mathbb{E}^x [\mathbf{X}_{0,t}^i] \right. \\ (3.5) \quad &+ \sum_{i,j=1}^d \frac{1}{2} W_i W_j f(z) \mathbb{E}^x [\mathbf{X}_{0,t}^i \mathbf{X}_{0,t}^{x;j}] \\ &+ \left. \sum_{i,j=1}^d \frac{1}{2} [W_i, W_j] f(z) \mathbb{E}^x [\mathbf{X}_{0,t}^{i,j}] \right) g(z) \mu(dz). \end{aligned}$$

Since  $p_1 : \mathcal{O} \mapsto G$  is a (surjective) submersion (in fact a smooth fibre bundle), by integrating over the fibres (e.g., [12] p.307) we can associate to any  $v \in C_c^\infty(U)$  a function  $\hat{v} \in C_c^\infty(G)$  such that for any bounded measurable  $h : G \mapsto \mathbb{R}$

$$\int_U (h \circ p_1)(z) v(z) \mu(dz) = \int_G h(x) \hat{v}(x) \lambda(dx).$$

In particular, setting  $v_i := (W_i f)g$ ,  $v_{i,j} := (W_i W_j f)g$  and  $w_{i,j} := ([W_i, W_j]f)g$ , we can apply Lemma 3.7 to obtain that (3.5) equals

$$(3.6) \quad \sum_{i,j=1}^d \int_G [-a^{i,j}(x)(U_j \hat{v}_i)(x) + a^{i,j}(x)\hat{v}_{i,j}(x)] \lambda(dx).$$

It remains to show that (3.6) agrees with the RHS of (3.4). To this end, we may assume by a limiting argument that  $a$  is smooth, and note that the same argument as [9] p.503 applies *mutatis mutandis* to our current setting.  $\square$

*Proof of Theorem 3.2.* Consider an increasing sequence of pre-compact open sets  $(U_n)_{n \geq 1}$  such that  $\cup_{n \geq 1} U_n = \mathcal{O}$ . By Lemma 3.8, we can apply Proposition 3.1 to conclude that for every  $x \in \mathcal{O}$  and  $n \geq 1$  such that  $x \in U_n$ , there exists a kernel  $p_t^n(x, \cdot) \in L^2(U_n, \mu)$  such that  $P_t^{U_n} f(x) = \langle p_t^n(x, \cdot), f \rangle_{L^2(U_n, \mu)}$  for all  $f \in C_b(U_n)$ . Moreover, by definition of  $P_t^{U_n}$ , the sequence  $p_t^n(x, \cdot)$  is increasing in  $n$  and satisfies  $\|p_t^n(x, \cdot)\|_{L^1(U_n, \mu)} \leq 1$ . Hence the limit  $p_t(x, \cdot) := \lim_{n \rightarrow \infty} p_t^n(x, \cdot)$  is almost everywhere finite and gives precisely the transition kernel of the Markov process  $\mathbf{Z}_t$  in  $\mathcal{O}$  with respect to  $\mu$ .  $\square$

#### APPENDIX A. PROOF OF PROPOSITION 3.1

We follow the notation from Section 3.1. For  $f \in C_c^\infty(U)$  denote

$$\|Wf\|_2^2 := \sum_{i=1}^d \|W_i f\|_2^2,$$

and for  $\alpha > 0$

$$\mathcal{E}_\alpha(f, f) := \mathcal{E}(f, f) + \alpha \|f\|_2^2.$$

**Lemma A.1.** (1) For every  $\varepsilon < \Lambda^{-1}$ , there exists  $\alpha > 0$ , depending only on  $\varepsilon$ ,  $\Lambda$ ,  $\|a\|_\infty$  and  $\sum_{i=1}^d \|\operatorname{div}_\mu W_i\|_\infty$ , such that for all  $f, g \in C_c^\infty(U)$

$$(A.1) \quad \mathcal{E}_\alpha(f, f) \geq \varepsilon \|Wf\|_2^2.$$

(2) There exist  $\beta > 0$ , depending only on  $\|a\|_\infty$  and  $\sum_{i=1}^d \|\operatorname{div}_\mu W_i\|_\infty$ , such that

$$(A.2) \quad |\mathcal{E}(f, g)| \leq \beta \|Wf\|_2 (\|Wg\|_2 + \|g\|_2).$$

*Proof.* By the Cauchy-Schwartz inequality and (3.1), for some  $C_1, \alpha > 0$

$$\begin{aligned} \mathcal{E}(f, f) &= \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) W_j f(z) \mu(dz) + \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) \operatorname{div}_\mu W_j(z) f(z) \mu(dz) \\ &\geq \sum_{i=1}^d \int_U \Lambda^{-1} |W_i f(z)|^2 \mu(dz) - \sum_{i,j=1}^d \|\operatorname{div}_\mu W_j\|_\infty \|a^{i,j}\|_\infty \|W_i f\|_2 \|f\|_2 \\ &\geq \Lambda^{-1} \|Wf\|_2^2 - C_1 \|Wf\|_2 \|f\|_2 \\ &\geq \varepsilon \|Wf\|_2^2 - \alpha \|f\|_2^2, \end{aligned}$$

which implies (A.1). On the other hand, by Cauchy-Schwartz, for some  $C_2, C_3 > 0$

$$\left| \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) W_j g(z) dz \right| \leq C_2 \|Wf\|_2 \|Wg\|_2,$$

and

$$\left| \sum_{i,j=1}^d \int_U a^{i,j}(z) W_i f(z) \operatorname{div}_\mu W_j(z) g(z) dz \right| \leq C_3 \|Wf\|_2 \|g\|_2,$$

from which we obtain (A.2).  $\square$

Since  $W$  satisfies Hörmander's condition on  $\mathcal{O}$ , recall that for every  $x \in \mathcal{O}$  there exist a neighbourhood  $U_x$  of  $x$  with  $\mu(U_x) < \infty$ ,  $\nu_x > 2$ , and  $C_x > 0$  such that for all  $f \in C_c^\infty(U_x)$  (e.g., [19] p.296)

$$\left( \int_{U_x} |f|^{2\nu_x/(\nu_x-2)} d\mu \right)^{(\nu_x-2)/\nu_x} \leq C_x \int_{U_x} \left( \sum_{i=1}^d |W_i f|^2 + |f|^2 \right) d\mu.$$

Since  $U$  is pre-compact, it is routine to patch together such inequalities using a partition of unity and apply interpolation to arrive at the following Sobolev inequality.

**Lemma A.2** (Sobolev inequality). *There exist constants  $\nu > 2$  and  $C, R > 0$  such that for all  $f \in C_c^\infty(U)$  with  $\|f\|_1 \leq 1$  and  $\|f\|_2 > R$*

$$\|f\|_{2\nu/(\nu-2)}^2 \leq C \|Wf\|_2^2.$$

Fix  $\varepsilon < \Lambda^{-1}$  and  $\alpha > 0$  such that (A.1) holds. Let  $\mathcal{F}$  be the closure of  $C_c^\infty(U)$  under  $\|\cdot\|_{\mathcal{F}} := \mathcal{E}_\alpha(\cdot, \cdot)^{1/2}$ .

**Corollary A.3** (Nash inequality). *Let  $\nu > 2$  and  $R > 0$  be the same as in Lemma A.2. There exists  $c > 0$  such that for all  $f \in \mathcal{F}$  with  $\|f\|_1 \leq 1$  and  $\|f\|_2 > R$ , it holds that*

$$(A.3) \quad \mathcal{E}(f, f) \geq c \|f\|_2^{2+4/\nu}.$$

*Proof.* Consider first  $f \in C_c^\infty(U)$ . The Sobolev inequality (Lemma A.2), along with Hölder's inequality, implies that

$$\|f\|_2^{2+4/\nu} \leq C \|f\|_1^{4/\nu} \|Wf\|_2^2,$$

from which the conclusion follows first for all  $f \in C_c^\infty(U)$  by (A.1), and then for general  $f \in \mathcal{F}$  by an approximation argument.  $\square$

*Proof of Proposition 3.1.* The desired properties of  $\mathcal{E}$  all follow from (A.1) and (A.2) and the fact that each  $W_i$  is a closable operator defined on  $C_c^\infty(U) \subset L^2(U, \mu)$ .

Denote by  $A$  the generator of the associated adjoint semi-group  $P_t^*$  in  $L^2(U, \mu)$  with domain  $D(A)$ . Consider  $f \in D(A)$  with  $\|f\|_1 \leq 1$  and set  $u_t = P_t^* f$ . Since  $P_t$  is sub-Markov, we have  $\|u_t\|_1 \leq 1$ , so by Corollary A.3, whenever  $\|u_t\|_2 > R$ ,

$$\frac{d}{dt} \|u_t\|_2^2 = \lim_{h \rightarrow 0} \frac{\|P_h^* u_t\|_2^2 - \|u_t\|_2^2}{h} = -2\mathcal{E}(u_t, u_t) \leq -2c \|u_t\|_2^{2+4/\nu},$$

from which it follows that there exists  $b > 0$  such that  $\|P_t^* f\|_2 \leq bt^{-\nu/2}$ .

To complete the proof, it remains only to apply an approximation of the Dirac delta  $\langle \phi, f_n \rangle \rightarrow \langle \phi, \delta_x \rangle = \phi(x)$  for all  $\phi \in C_b(U)$ , with  $f_n \in D(A)$  and  $\|f_1\| \leq 1$ , and use the fact that  $\sup_n \|P_t^* f_n\|_2 \leq bt^{-\nu/2}$  and that  $P_t$  preserves  $C_b(U)$ .  $\square$

## REFERENCES

- [1] Andrei A. Agrachev and Yuri L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.
- [2] R. F. Bass, B. M. Hambly, and T. J. Lyons. Extending the Wong-Zakai theorem to reversible Markov processes. *J. Eur. Math. Soc. (JEMS)*, 4(3):237–269, 2002.
- [3] Fabrice Baudoin and Martin Hairer. A version of Hörmander’s theorem for the fractional Brownian motion. *Probab. Theory Related Fields*, 139(3-4):373–395, 2007.
- [4] Thomas Cass and Peter Friz. Densities for rough differential equations under Hörmander’s condition. *Ann. of Math. (2)*, 171(3):2115–2141, 2010.
- [5] Thomas Cass, Martin Hairer, Christian Litterer, and Samy Tindel. Smoothness of the density for solutions to Gaussian rough differential equations. *Ann. Probab.*, 43(1):188–239, 2015.
- [6] Thomas Cass and Marcel Ogrodnik. Tail estimates for markovian rough paths. arXiv:1411.5189, May 2015. To appear in Annals of Probability.
- [7] Ilya Chevyrev and Terry Lyons. Characteristic functions of measures on geometric rough paths. *Ann. Probab.*, 44(6):4049–4082, 2016.
- [8] P. J. Fitzsimmons. Absolute continuity of symmetric diffusions. *Ann. Probab.*, 25(1):230–258, 1997.
- [9] Peter Friz and Nicolas Victoir. On uniformly subelliptic operators and stochastic area. *Probab. Theory Related Fields*, 142(3-4):475–523, 2008.
- [10] Peter K. Friz, Benjamin Gess, Archil Gulisashvili, and Sebastian Riedel. The Jain–Monrad criterion for rough paths and applications to random Fourier series and non-Markovian Hörmander theory. *Ann. Probab.*, 44(1):684–738, 2016.
- [11] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
- [12] Victor Guillemin and Shlomo Sternberg. *Geometric asymptotics*. American Mathematical Society, Providence, R.I., 1977. Mathematical Surveys, No. 14.
- [13] Antoine Lejay. Stochastic differential equations driven by processes generated by divergence form operators. I. A Wong-Zakai theorem. *ESAIM Probab. Stat.*, 10:356–379, 2006.
- [14] Antoine Lejay. Stochastic differential equations driven by processes generated by divergence form operators. II. Convergence results. *ESAIM Probab. Stat.*, 12:387–411, 2008.
- [15] Terry Lyons and Lucrețiu Stoica. The limits of stochastic integrals of differential forms. *Ann. Probab.*, 27(1):1–49, 1999.
- [16] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.
- [17] J. Nash. Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.*, 80:931–954, 1958.
- [18] Yoichi Oshima. *Semi-Dirichlet forms and Markov processes*, volume 48 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2013.
- [19] Karl-Theodor Sturm. Analysis on local Dirichlet spaces. II. Upper Gaussian estimates for the fundamental solutions of parabolic equations. *Osaka J. Math.*, 32(2):275–312, 1995.

I. CHEVYREV, MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, ANDREW WILES BUILDING, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, UNITED KINGDOM

*E-mail address:* chevyrev@maths.ox.ac.uk

M. OGRODNIK, DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON, HUXLEY BUILDING, 180 QUEENS GATE, LONDON SW7 2AZ, UNITED KINGDOM

*E-mail address:* m.ogrodnik12@imperial.ac.uk